

Modelos de elección discreta para datos de panel y modelos de duración: una revisión de la literatura

Raquel Carrasco Perea
Universidad Carlos III de Madrid

Octubre 2001

Resumen

En este artículo se revisa la literatura sobre los modelos de elección discreta para datos de panel y los modelos de duración en tiempo discreto. Respecto a los primeros, se analizan los modelos de efectos fijos y de efectos aleatorios, así como los problemas de correlación serial y dependencia de los estados en modelos dinámicos. Se presentan diferentes estimadores que permiten eliminar los sesgos debidos a la existencia de heterogeneidad inobservable entre individuos. Respecto a los modelos de duración en tiempo discreto, se estudia cómo éstos pueden estimarse como una secuencia de modelos de elección discreta, poniéndose de manifiesto la estrecha relación entre ambos tipos de modelos. También se destaca la posibilidad de eliminar la heterogeneidad inobservable utilizando las técnicas de efectos fijos y efectos aleatorios, de forma similar a como se hace en modelos de elección discreta para datos de panel. Para ello es necesario contar con paneles de datos de duración. Finalmente, se considera la cuestión de en qué casos es más apropiado utilizar modelos de elección discreta para datos de panel o modelos de duración, dependiendo de la naturaleza de los datos disponibles.

Palabras clave: modelo econométrico, modelo de elección discreta, modelo de duración, modelo dinámico, modelo de panel.

Clasificación JEL: C15, C23, C29.

Abstract

This article contains a review of the literature on binary choice panel data and discrete duration models. The discussion of the former addresses the problems arising in connection with fixed vs random effects, as well as with serial correlation and state dependence in dynamic models, focusing on the different estimators used to eliminate bias due to the existence of unmeasured heterogeneity. The study of discrete duration data models explores how they may be estimated as a sequence of binary choice models and draws attention to the close relationship between the two types of models. The possibility of eliminating unmeasured heterogeneity is also considered, by using fixed and random effect techniques in much the same way as in binary choice panel data models. Finally, the paper analyses when binary choice panel data are more appropriate than discrete duration models and vice versa, depending on the nature of the data.

Keywords: econometric model, discrete duration models, duration model, dynamic model, panel model.

Jel Classification: C15, C23, C29.

1. Introducción

Los datos de panel son ampliamente utilizados en microeconometría. La ventaja fundamental de un panel de datos es que nos permite estudiar cuestiones que no pueden ser abordadas en el contexto de sección cruzada o de series temporales, únicamente. En particular, los datos de panel permiten tener en cuenta la existencia de efectos individuales inobservables, que pueden estar correlacionados con otras variables incluidas en la especificación de una relación econométrica. Otra ventaja fundamental de los datos de panel es que permiten modelizar relaciones dinámicas entre las variables. En este contexto, la distinción entre variables estrictamente exógenas y predeterminadas se considera fundamental en el análisis aplicado.

En la literatura de datos de panel hay una gran cantidad de trabajos empíricos que se han concentrado en estudiar estas cuestiones en el ámbito de modelos lineales. Sin embargo, hay muchos menos resultados disponibles para modelos no lineales de elección discreta. En este caso, la presencia de efectos individuales inobservables y la consideración de relaciones dinámicas entre la variable dependiente y las variables explicativas genera nuevos problemas que no aparecen en el caso de los modelos lineales.

En este artículo se revisan los métodos utilizados en la estimación de modelos de elección binaria para datos de panel. La atención se centra en presentar diferentes modelos y estimadores que permiten eliminar el sesgo debido a la existencia de heterogeneidad inobservable e invariante en el tiempo entre individuos y de tener en cuenta la existencia de variables explicativas predeterminadas. Se estudia el llamado enfoque de «efectos-fijos». En principio este enfoque es muy atractivo, ya que no requiere realizar supuestos paramétricos sobre la distribución condicional de la heterogeneidad inobservable dadas las variables explicativas. Sin embargo, una desventaja de este método es que sólo puede utilizarse para ciertas distribuciones y, normalmente, requiere hacer supuestos muy restrictivos sobre la distribución de los errores. Estos problemas se pueden resolver, en parte, mediante el enfoque de «efectos aleatorios». Aunque este enfoque siempre puede utilizarse, las características restrictivas del enfoque anterior se eliminan a costa de imponer una especificación paramétrica para la distribución condicional de los efectos individuales inobservables.

En el artículo se discuten también los llamados «modelos de duración». Aunque el análisis de datos de duración es bastante reciente, constituye un área de creciente interés en econometría. Uno de los procesos más ampliamente estudiados en este contexto es el de la duración de los procesos de desempleo de los individuos. La investigación reciente demuestra que cuanto mayor es la duración del periodo de desempleo de un individuo, más probable es que éste se encuentre desempleado en un momento del tiempo. En este caso, el interés no se centra únicamente en la duración del proceso en sí, sino también en la probabilidad de que dicho proceso termine en el «siguiente periodo», dado que aún no ha terminado. Estas cuestiones pueden analizarse mediante la formulación y estimación de modelos de duración. En este artículo se plantean las cuestiones fundamentales que surgen en el estudio de datos de duración en tiempo discreto. Se analiza cómo estos modelos pueden estimarse como una secuencia de modelos de elección discreta, destacándose la estrecha relación existente entre ambos. En primer lugar, se presentan modelos de duración para datos en los que se observa una única duración de los individuos en un único estado. A continuación se consideran modelos de múltiples duraciones y múltiples estados. Se destaca que la existencia de datos de panel para duraciones permite separar el efecto de la dependencia de la duración del efecto de la heterogeneidad inobservable, de forma similar a como se hace en modelos de elección discreta para datos de panel.

Finalmente, también se considera la cuestión de en qué casos es más apropiado utilizar modelos de elección binaria o modelos de duración, dependiendo de la naturaleza de los datos disponibles. Los modelos de elección discreta para datos de panel son adecuados cuando el periodo observacional se corresponde con decisiones individuales. En muchos casos, sin embargo, los datos únicamente ofrecen información sobre una secuencia de esta-



dos junto con el tiempo que el individuo permanece en cada estado. En ese caso, sería más apropiado especificar modelos de duración.

La organización del artículo es la siguiente. La segunda sección discute los modelos de elección discreta para datos de panel. Esta sección se centra en el estudio de los modelos de efectos fijos y de efectos aleatorios. Además, se consideran los problemas de las condiciones iniciales y de la distinción entre correlación serial y dependencia de los estados en modelos dinámicos. A lo largo de esta sección se utiliza el caso bien conocido de la regresión lineal como guía de la exposición. La tercera sección examina los modelos de duración. Se estudian los datos de duración en el contexto de modelos discretos, destacando la estrecha correspondencia entre modelos binarios y modelos de duración en tiempo discreto. La sección cuarta se centra en las diferentes interpretaciones de una secuencia binaria, dependiendo de la naturaleza del intervalo muestral. Finalmente, la quinta sección presenta algunas conclusiones y apunta algunas líneas de investigación futura.

2. Modelos de elección binaria

En esta sección se consideran modelos en los que la variable dependiente y_i puede tomar sólo dos valores, que por conveniencia y sin pérdida de generalidad, se denotan por 1 si el suceso ocurre y 0 si no. Por ejemplo, y_i puede definirse como 1 si el individuo está en el mercado de trabajo (0 en caso contrario), o como 1 si el individuo pertenece a un sindicato (0 en caso contrario).

En algunas ocasiones es conveniente modelizar las variables discretas en términos de variables continuas latentes. Sea y_i igual a 0 – 1 donde:

$$y_i = \begin{cases} 1 & \text{si } y_i^* \geq 0, \\ 0 & \text{si } y_i^* < 0, \end{cases} \quad [1]$$

y la variable latente y_i^* es una función lineal de x_i :

$$y_i^* = \beta' x_i + u_i. \quad [2]$$

Por tanto, en estos modelos la variable dependiente y_i refleja decisiones de los individuos. Por ello, parece natural modelizar la probabilidad de un suceso condicional a un conjunto de variables exógenas:

$$E(y_i | x_i) = Pr(y_i = 1 | x_i) = F(\beta' x_i), \quad [3]$$

donde $F(\cdot)$ es la función de distribución acumulada (fda.). En las aplicaciones empíricas se han utilizado diversas formas funcionales para F . Por ejemplo, si se asume una función de regresión lineal se obtiene el llamado «modelo lineal de probabilidad»:

$$E(y_i | x_i) = Pr(y_i = 1 | x_i) = \beta' x_i. \quad [4]$$

Este modelo es muy sencillo y puede estimarse por el método de los mínimos cuadrados. Sin embargo, la principal objeción es que en muchos casos la esperanza condicional $E(y_i | x_i)$ no puede interpretarse como la probabilidad de que el suceso ocurra, ya que dicha esperanza puede estar fuera del intervalo (0, 1). Por ello F se suele especificar como una distribución logística (modelo Logit):

$$F(z) = \frac{e^z}{1 + e^z}, \quad [5]$$

o como una distribución normal $N(0, 1)$ (modelo Probit):

$$F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du, \quad [6]$$

ya que, dado que son funciones de distribución, están acotadas entre 0 y 1. En ambos casos se obtienen modelos de regresión no lineales que pueden estimarse por el método de máxima verosimilitud (MV). Para una muestra aleatoria de N individuos la función de verosimilitud de estos tres modelos se puede escribir de forma general como:

$$L = \prod_{i=1}^N F(\beta'x_i)^{y_i} (1 - F(\beta'x_i))^{1-y_i}. \quad [7]$$

Con datos de panel es posible relajar y contrastar algunos de los supuestos implícitos en el contexto de sección cruzada. En concreto, es posible tener en cuenta la existencia de heterogeneidad inobservable e invariante en el tiempo entre individuos. Así, añadiendo un efecto individual a la anterior especificación, podemos considerar el siguiente modelo de elección discreta para N individuos observados durante T periodos consecutivos:

$$y_{it}^* = \beta'x_{it} + \eta_i + u_{it} \quad (i = 1, 2, \dots, N, t = 1, 2, \dots, T), \quad [8]$$

$$Pr(y_{it} = 1 | x_i^T, \eta_i) = F(\beta'x_{it} + \eta_i). \quad [9]$$

Para estimar este modelo, η_i puede tratarse como un conjunto adicional de parámetros a estimar junto con β . Este es el llamado enfoque de «efectos fijos». Sin embargo este método da lugar a estimaciones inconsistentes de los parámetros de interés, β , cuando N es grande pero T es fijo, debido al problema de los parámetros incidentales. Una solución, dentro del enfoque de efectos fijos, consiste en maximizar una función de verosimilitud condicional a estadísticos suficientes para los parámetros incidentales η_i . Sin embargo, este enfoque sólo es válido para algunas especificaciones de $F(\cdot)$. También es posible utilizar un enfoque de efectos aleatorios, que implica asumir una distribución para los parámetros incidentales. El problema es que en algunas ocasiones los supuestos requeridos para aplicar este método son muy restrictivos. Estas cuestiones se analizan a continuación.

2.1. El modelo de efectos fijos

Repasemos brevemente el caso de la regresión lineal. Consideremos el siguiente modelo lineal para datos de panel:

$$y_{it} = \beta'x_{it} + \eta_i + u_{it} \quad (i = 1, 2, \dots, N, t = 1, 2, \dots, T) \quad [10]$$

donde u_{it} se distribuye $iid N(0, \sigma^2)$. Los efectos fijos η_i se pueden tratar como un conjunto adicional de N coeficientes que se estiman conjuntamente con β mediante, por ejemplo, una regresión múltiple de y sobre x y un conjunto de variables indicadoras. Este es el llamado enfoque de efectos fijos, frente al enfoque de efectos aleatorios. La principal diferencia entre ambos es que, mientras que en el contexto de efectos aleatorios es necesario asumir una distribución para η_i , bajo el enfoque de efectos fijos se condiciona en η_i , de forma que su distribución no representa ningún papel. Una simplificación computacional muy útil en este caso consiste en obtener el estimador por MV de β a partir de una regresión de $y_{it} - \bar{y}_i$ sobre $x_{it} - \bar{x}_i$, donde \bar{y}_i y \bar{x}_i son medias individuales ($\bar{y}_i = \sum_{t=1}^T y_{it}$). En el caso de $T = 2$ este enfoque es equivalente a realizar una regresión de $y_{i2} - y_{i1}$ sobre $x_{i2} - x_{i1}$. Dado que:

$$y_{i2} - y_{i1} = \beta(x_{i2} - x_{i1}) + (u_{i2} - u_{i1}) \quad [11]$$

con u independiente de x , esta regresión proporciona estimadores consistentes de β a medida que $N \rightarrow \infty$ (siempre que haya suficiente variación en $x_2 - x_1$).

En el caso de modelos de elección discreta no lineales, el problema de encontrar un estimador consistente es más complicado. Consideremos el siguiente modelo de probabilidad:

$$Pr(y_{it} = 1 \mid x_i^T, \eta_i) = F(\beta'x_{it} + \eta_i), \quad [12]$$

donde $x_i^T = (x_{i1}, \dots, x_{iT})$.

A diferencia del modelo lineal de probabilidad, tomar primeras diferencias en este caso no elimina los efectos individuales. Por tanto, para estimar este modelo los η_i podrían tratarse como parámetros adicionales. Sin embargo, este método produce estimaciones inconsistentes de los parámetros de interés β cuando N es grande pero T es fijo¹, debido al problema de los parámetros incidentales. La idea es que, al contrario que en el caso de la regresión lineal, los estimadores de MV para η_i y β no son independientes entre sí. Si T es fijo, los estimadores por MV de los η_i son inconsistentes. Debido a la no linealidad del modelo, β ha de estimarse conjuntamente con los η_i y la inconsistencia de $\hat{\eta}_i$ se transmite a $\hat{\beta}$. Por tanto, incluso si N tiende a infinito, el estimador por MV de β es inconsistente, a diferencia del caso de regresión lineal donde es posible encontrar un estimador de β que no dependa del efecto fijo estimado.

Andersen (1973) y Chamberlain (1980) demostraron esto para el modelo Logit y sugirieron utilizar un enfoque de verosimilitud condicional. La ventaja de este enfoque consiste en que no se imponen restricciones sobre la distribución condicional de η dado x ; aunque

¹ Dado que los paneles microeconómicos típicos contienen un gran número de individuos observados en un periodo corto de tiempo, las distribuciones asintóticas relevantes son aquellas para las que el número de individuos crece, pero la dimensión temporal es fija.

también tiene algunas desventajas. Un supuesto clave que se requiere es que las variables explicativas sean estrictamente exógenas. La idea consiste en basar la función de verosimilitud en la distribución de los datos condicional a un conjunto de estadísticos suficientes para los parámetros incidentales. Como se discutió anteriormente, en el caso de regresión lineal un estadístico suficiente para η_i es Σy_{it} , y la función de verosimilitud condicional se corresponde con el modelo en desviaciones respecto a las medias individuales.

En el caso del modelo Logit, el enfoque de verosimilitud condicional se puede aplicar directamente. Sea:

$$Pr(y_{it} | x_i^T, \eta_i) = F(\beta' x_{it} + \eta_i), \quad F(z) = \frac{e^z}{1 + e^z}, \quad [13]$$

donde y_{i1}, \dots, y_{iT} son independientes condicionales en x_{i1}, \dots, x_{iT} , η_i . En este caso, $\Sigma_t y_{it}$ es de nuevo un estadístico suficiente para η_i . Consideremos el caso de $T = 2$. Si $y_{i1} + y_{i2} = 0$ ó 2 , entonces y_{i1} y y_{i2} quedan determinados por su suma. Por tanto, el único caso de interés es aquél en que $y_{i1} + y_{i2} = 1$. Así, las dos posibilidades que se consideran son:

$$w_i = \begin{cases} 1, & \text{si } (y_{i1}, y_{i2}) = (0, 1), \\ 0, & \text{si } (y_{i1}, y_{i2}) = (1, 0). \end{cases}$$

La densidad condicional tiene la siguiente forma:

$$Pr(y_{i2} = 1 | x_{i1}, x_{i2}, \eta_i, y_{i1} + y_{i2} = 1) = F[\beta'(x_{i2} - x_{i1})], \quad [14]$$

que no depende de η_i . Dada una muestra aleatoria de individuos, la función de verosimilitud condicional (en logaritmos) es:

$$L = \sum_{i \in B} \{w_i \ln F[\beta'(x_{i2} - x_{i1})] + (1 - w_i) \ln F[-\beta'(x_{i2} - x_{i1})]\},$$

donde

$$B = \{i | y_{i1} + y_{i2} = 1\}.$$

Es importante destacar que, utilizando este enfoque, aquellos individuos para los que $\Sigma_t y_{it} = 0$ ó $\Sigma_t y_{it} = T$, es decir, individuos que nunca cambian de estado, no son utilizados en la estimación ya que su contribución a la verosimilitud es nula.

Esta función de verosimilitud condicional no depende de los parámetros incidentales η_i y tiene la forma de una función de verosimilitud de un modelo Logit, en el que los dos resultados posibles son (0, 1) y (1, 0), y la variable explicativa es $x_{i2} - x_{i1}$. Chamberlain (1980) demostró que este método puede extenderse al caso Logit multinomial y al modelo log-lineal. Sin embargo, en el modelo Probit no es posible encontrar un estadístico suficiente para η_i .

El enfoque de verosimilitud condicional puede considerarse «análogo» a la regresión

en primeras diferencias en el modelo lineal. Sin embargo, a diferencia del modelo lineal, los estimadores en este caso sólo son consistentes si las perturbaciones para cada individuo son ruido blanco con distribución logística y si las variables explicativas varían en el tiempo. No obstante, Manski (1987) demostró que es posible relajar el supuesto de distribución logística y que basta con asumir que las perturbaciones son estacionarias con soporte no acotado y que las variables explicativas tienen suficiente variación temporal. Manski probó que una versión condicional del estimador de máximo *score* (Manski, 1985) proporciona estimaciones consistentes de los parámetros. Este tipo de modelos tiene una interpretación como regresión mediana:

$$\text{med}(y_{i2} - y_{i1} | x_i, y_{i1} + y_{i2} = 1) = \text{sgn}(\beta'(x_{i2} - x_{i1})), \quad [15]$$

donde $\text{med}(\cdot)$ es la mediana y $\text{sgn}(\cdot)$ es una función definida como $\text{sgn}(u) = -1$ si $u < 0$, $\text{sgn}(u) = 0$ si $u = 0$ y $\text{sgn}(u) = 1$ si $u > 0$. Manski probó que el vector de parámetros normalizados $\beta^* = \frac{\beta}{\|\beta\|}$ está identificado siempre que las variables explicativas varíen suficientemente en el tiempo. Propuso estimar β^* mediante la maximización de la siguiente función:

$$s(\beta) = \frac{1}{N} \sum_{i=1}^N (y_{i2} - y_{i1}) \text{sgn}(\beta'(x_{i2} - x_{i1})). \quad [16]$$

Por tanto, el estimador que maximiza $s(\beta)$ es un estimador de máximo *score* aplicado a las primeras diferencias de los datos en la submuestra para la que $y_{i1} \neq y_{i2}$. Este estimador es consistente, aunque su tasa de convergencia es menor que \sqrt{N} . Por ello no es posible obtener su distribución asintótica utilizando la teoría asintótica estándar. Otro problema es que la función $s(\beta)$ tendrá, en general, varios máximos locales, por lo que la elección de los valores iniciales es crucial.

En esta línea, Kyriazidou (1997) propuso estimadores para modelos de selección para datos de panel, en los que tanto la ecuación de selección, como la de regresión, contienen efectos individuales inobservables. La ecuación de regresión también puede contener retardos de la variable dependiente continua, mientras que la ecuación de selección puede contener valores retardados de la variable indicador. Kyriazidou presenta un estimador en dos etapas. En la primera etapa, los coeficientes de la ecuación de selección se estiman consistentemente utilizando una versión «suavizada» del estimador de máximo *score* condicional de Manski (1987). Sus simulaciones de Monte Carlo muestran que este estimador puede funcionar bien en la práctica, con bases de datos suficientemente grandes. Sin embargo, es bastante sensible a la elección del parámetro de ancho de banda.

2.1.1. Modelos de elección binaria dinámicos

En los modelos estáticos discutidos en la sección anterior se supone que la probabilidad de cambiar de estado es independiente de que el suceso haya ocurrido o no en el pasado. Sin embargo, en una gran variedad de contextos, como, por ejemplo, en el estudio de la participación en el mercado de trabajo, decisiones de compra, etcétera, a menudo ocurre

que los individuos que han experimentado un suceso en el pasado tienen una mayor probabilidad de experimentarlo en el futuro. Heckman (1981) analiza dos explicaciones para este fenómeno. La primera es la llamada «dependencia de los estados verdadera», en el sentido de que la decisión pasada entra en el modelo de una forma estructural como una variable explicativa. La segunda es la existencia de correlación serial en las perturbaciones del modelo. De particular interés es el caso en el que esta correlación serial se debe a la presencia de heterogeneidad inobservable individual. Heckman llama a esta fuente de correlación serial «dependencia de los estados espúrea».

Distinguir entre estas dos explicaciones es importante, por ejemplo, en el análisis de las secuencias de empleo y desempleo de un individuo. En ese caso, una cuestión básica consiste en analizar si el desempleo actual es causa del desempleo futuro o no (Heckman, 1981; Card y Sullivan, 1988). Heckman señala que para poder discriminar entre dependencia de los estados verdadera o espúrea es necesario disponer de datos de panel del historial de los individuos.

Otra cuestión que surge en el contexto de los modelos dinámicos es el problema de las condiciones iniciales, fundamental para obtener estimadores consistentes. Normalmente, el inicio del periodo muestral no suele coincidir con el inicio del proceso estocástico que genera las observaciones. Por tanto, al estimar modelos dinámicos es necesario asumir el comportamiento de los valores iniciales y_{i0} . De hecho, distintos supuestos sobre las condiciones iniciales dan lugar a funciones de verosimilitud diferentes². Por tanto, elecciones erróneas de las condiciones iniciales pueden dar lugar a estimadores inconsistentes. En la mayoría del trabajo aplicado se han hecho dos supuestos sobre las condiciones iniciales: que son exógenas o que el proceso está en equilibrio.

Consideremos el siguiente modelo en términos de una variable latente:

$$y_{it}^* = \gamma_{it-1} + \beta' x_{it} + \eta_i + u_{it}, \quad [17]$$

donde $y_{it} = 1$ si $y_{it}^* > 0$, 0 en caso contrario.

Bajo el supuesto de que y_{i0} es una constante fija para el individuo i , la probabilidad conjunta de $y_i' = (y_{i1}, \dots, y_{iT})$ dado η_i es:

$$\prod_{t=1}^T F(y_{it} | y_{it-1}, \eta_i) = \prod_{t=1}^T F(\gamma_{it-1} + \beta' x_{it} + \eta_i)^{y_{it}} (1 - F(\gamma_{it-1} + \beta' x_{it} + \eta_i))^{1-y_{it}}. \quad [18]$$

Bajo el supuesto de que el proceso está en equilibrio, el límite de la probabilidad marginal de $y_{it} = 1$ para todo t , dado η_i , es:

$$P_i = \frac{F(\beta' x_{it} + \eta_i)}{1 - F(\gamma + \beta' x_{it} + \eta_i) + F(\beta' x_{it} + \eta_i)}, \quad [19]$$

² Véase HSIAO (1986).

y el límite de la probabilidad de $y_{it} = 0$ es $1 - P_i$. Por tanto, la probabilidad conjunta de (y_{i0}, \dots, y_{iT}) , dado η_i , es:

$$\prod_{t=1}^T F(\gamma y_{it-1} + \beta' x_{it} + \eta_i)^{y_{it}} (1 - F(\gamma y_{it-1} + \beta' x_{it} + \eta_i))^{1-y_{it}} P_i^{y_{i0}} (1 - P_i)^{1-y_{i0}}. \quad [20]$$

En el caso de modelos lineales dinámicos, es bien conocido que la regresión en primeras diferencias, o en desviaciones respecto a la media, da lugar a estimadores inconsistentes. Por ello, los modelos lineales se suelen estimar utilizando el Método Generalizado de Momentos (MGM)³. Sin embargo, en el caso de los modelos no lineales tanto la identificación como la estimación presentan mayores dificultades.

El enfoque de efectos fijos se puede también utilizar para estimar modelos Logit para datos de panel con efectos individuales y con retardos de la variable dependiente, siempre que no haya otras variables explicativas en el modelo y que haya, al menos, cuatro observaciones por individuo (véase Chamberlain, 1985; Magnac, 1997). El modelo es el siguiente:

$$Pr(y_{i0} = 1 | \eta_i) = P_0(\eta_i),$$

$$Pr(y_{it} = 1 | \eta_i, y_{i0}, \dots, y_{it-1}) = \frac{\exp(\gamma y_{it-1} + \eta_i)}{1 + \exp(\gamma y_{it-1} + \eta_i)} \quad t = 1, \dots, T; T \geq 3, \quad [21]$$

donde se asume que y_{i0} se observa. Cuando $T = 3$, la inferencia sobre γ se basa en que la probabilidad

$$Pr(y_{i0} = d_0, y_{i1} = 0, y_{i2} = 1, y_{i3} = d_3 | y_{i1} + y_{i2} = 1, \eta_i),$$

es independiente de η_i . Este método tiene dos limitaciones importantes. La primera es que implica eliminar un gran número de observaciones y, por tanto, utiliza únicamente una pequeña proporción de los datos. El segundo problema es que el modelo no permite incluir variables exógenas, por lo que su uso para el análisis en términos de evaluación de políticas es limitado.

Honoré y Kyriazidou (1998) estudiaron la identificación y estimación de estos modelos cuando entre las variables explicativas se incluyen variables estrictamente exógenas, retardos de la variable dependiente y efectos individuales inobservables. Demostraron que tanto β como γ están identificados si se dispone de cuatro o más observaciones por individuo. Suponen que se observa y_{i0} , pero no necesitan asumir que las variables explicativas se observan en el periodo inicial. La identificación de este modelo se basa en derivar un conjunto de probabilidades independientes del efecto individual. Proponen un estimador que es consistente y asintóticamente normal, aunque la tasa de convergencia es más lenta que la inversa de la raíz cuadrada del tamaño muestral. El enfoque de Honoré-Kyriazidou presenta varios problemas. En primer lugar, se requiere estacionariedad, lo que excluye la posibi-

³ Modelos lineales autoregresivos o modelos lineales con variables predeterminadas han sido estudiados por ANDERSON y HSIAO (1981), HOLTZ-EAKING, NEWWEY y ROSEN (1988), ARELLANO y BOND (1991), y ARELLANO y BOVER (1995), entre otros.

lidad de incluir *dummies* temporales o de tener en cuenta heteroscedasticidad. Además, los errores han de distribuirse *iid* según una distribución logística e independiente de (x_i, η_i, y_{i0}) en todos los periodos, por lo que tampoco se permite correlación serial. Además, la identificación se basa en restringir la variación de x en el tiempo (por ejemplo, con $T = 3$ la inferencia se basa únicamente en observaciones para las que x_{i2} es «cercana» a x_{i3}). Finalmente, se requiere que todas las variables x 's sean estrictamente exógenas.

2.2. Modelos de efectos aleatorios: componentes del error y sesgo de heterogeneidad

Tradicionalmente la literatura de datos de panel ha tenido dos objetivos diferentes a la hora de analizar modelos en los que los efectos individuales son tratados como variables aleatorias: el análisis dinámico de los componentes del error y la eliminación del sesgo debido a la heterogeneidad inobservable.

2.2.1. Análisis de los componentes del error

Empecemos considerando el caso en el que la ecuación [8] representa un modelo binario de componentes del error, donde η_i se considera una variable aleatoria⁴. En este contexto, se asume que las variables explicativas x_{it} son estrictamente exógenas con respecto al error compuesto $v_{it} = \eta_i + u_{it}$. Por tanto, se supone que los efectos individuales η_i son independientes de x_i . La principal motivación en este caso consiste en separar los componentes de la varianza para medir persistencia. Por tanto, el interés se centra en estimar separadamente la variación permanente y transitoria en la variable y_{it} , en lugar de en estimar consistentemente el coeficiente de regresión β .

Si la secuencia de componentes inobservables $v_{i1}, v_{i2}, \dots, v_{iT}$ se trata como una secuencia de funciones indicador aleatorias, el modelo especifica la siguiente probabilidad:

$$Pr(y_{i1}^* = 1, \dots, y_{iT}^* = 1) = Pr(\beta'x_{i1} > v_{i1}, \dots, \beta'x_{iT} > v_{iT}). \quad [22]$$

El modelo se completa haciendo algún supuesto sobre la distribución de $v_i = (v_{i1}, \dots, v_{iT})'$. Si los v_{it} son independientes, los datos de panel son innecesarios. Sin embargo, este supuesto no es muy adecuado. Un supuesto muy flexible es el de que los errores se distribuyen según una distribución normal multivariante $N(0, \Omega)$. El problema es que para obtener estimaciones por MV es necesario evaluar integrales de dimensión T , lo que es computacionalmente intratable para T mayor que 3 ó 4. De nuevo, para poder realizar la estimación se suele asumir que los errores son *iid* o que están equicorrelacionados. Por ejemplo, en el caso de $\Omega = \sigma^2 I + \sigma_u^2 II'$ la integral múltiple puede escribirse como una integral univariante (véase Heckman, 1981; Butler y Moffitt, 1982), lo que simplifica enormemente la estimación.

Sin embargo, en muchos casos no es deseable utilizar estos supuestos simplificadores sobre los errores. Los métodos de estimación por simulación recientemente desarrollados

⁴ En el caso lineal, el modelo tradicional de componentes del error ha sido estudiado por BALESTRA y NERLOVE (1996).

proporcionan una forma muy flexible de aproximar las probabilidades sin necesidad de realizar este tipo de supuestos (véase Hajivassilou y McFadden, 1990; Hajivassilou, 1993, y Keane, 1994). Estos métodos se basan en el método de los momentos simulados (MMS) desarrollado por McFadden (1989) para la estimación de modelos Probits multinomiales para datos de sección cruzada. La idea consiste en realizar simulaciones de Monte Carlo de las integrales múltiples, en lugar de evaluarlas numéricamente. El estimador del MMS proporciona un método potente para estimar modelos de elección discreta con sendas muy complejas de correlación serial. Keane (1994) desarrolló una extensión del método de McFadden para el caso de datos de panel. Su estimador se basa en expresar la función objetivo en términos de probabilidades de transición (este método se conoce como Geweke-Hajivassilou-Keane, o simulador de GHK, ya que simultánea e independientemente fue desarrollado por Geweke y Hajivassilou).

Sin embargo, el interés de los modelos con efectos individuales independientes de las variables explicativas es escaso. Hay razones para creer que en muchas ocasiones η_i y x_{it} están correlacionados. Dada la importancia de reconocer esta posible correlación, parece más interesante considerar modelos de efectos aleatorios correlacionados con las variables explicativas.

2.2.2. Sesgos debidos a la heterogeneidad inobservable

El modelo de componentes del error tradicional asume que η es independiente de x . Sin embargo, el principal interés se centra en los casos en los que las variables inobservables están correlacionadas con x . Si erróneamente se modeliza η como independiente de x , entonces el sesgo por variables omitidas no se elimina. En este contexto el principal objetivo ha sido obtener estimadores que converjan a β a medida que el número de individuos crece, incluso si el número de periodos es pequeño. Una posibilidad sugerida por Chamberlain (1984) ha sido suponer que la dependencia entre η y x se produce a través de un función de regresión lineal:

$$\eta_i = \lambda' x_i^T + \xi_i, \quad [23]$$

donde ξ_i es independiente de las x 's y sigue una distribución normal $N(0, \sigma_\xi^2)$. Esta especificación conduce al siguiente modelo de efectos aleatorios:

$$\begin{aligned} y_{it}^* &= \beta' x_{it} + \lambda' x_i^T + \xi_i + u_{it} \\ &= \pi_1 x_{i1} + \dots + \pi_T x_{iT} + \varepsilon_{it}, \end{aligned} \quad [24]$$

donde $\pi_s = \beta + \lambda_s$ para $t = s$, $\pi_s = \lambda_s$ para $t \neq s$ y $(\varepsilon_{i1}, \dots, \varepsilon_{iT})$ sigue una distribución normal conjunta. La estimación eficiente de este modelo se enfrenta a los mismos problemas computacionales que la estimación de los modelos de componentes del error. Sin embargo, es posible estimar β y λ utilizando un estimador de distancia mínima. Chamberlain sugirió estimar cada una de los t probits univariantes de sección cruzada separadamente por MV. Para ello, definamos $\hat{\pi} = [\hat{\pi}_1 \dots \hat{\pi}_T]$, donde $\hat{\pi}_t$ es el estimador de MV de π_t en la t -ésima ecuación [24]. Dado que β y λ son funciones de π , pueden estimarse conjuntamente minimizando la distancia cuadrática entre $\hat{\pi}$ y $\pi(\beta, \lambda)$:

$$d(\beta, \lambda) = [\hat{\pi} - \pi(\beta, \lambda)]' V^{-1} [\hat{\pi} - \pi(\beta, \lambda)]. \quad [25]$$

La elección óptima de V es un estimador consistente de la varianza asintótica de $\hat{\pi}$.

No obstante, la minimización de [25] presenta varios problemas. Uno de ellos es que es necesario estimar conjuntamente β y λ , a pesar de que únicamente estemos interesados en β . Además, para evaluar [25] se requieren procedimientos iterativos, ya que en muchas ocasiones las restricciones en π son no lineales. Un procedimiento alternativo para estimar este tipo de modelos de efectos aleatorios es el propuesto por Bover y Arellano (1988). Sugieren calcular predicciones para las variables latentes y_{it}^* utilizando la forma reducida [24]: $\hat{y}_{it}^* = \hat{\pi}_i^* x_{it}^T$, y después estimar β mediante una regresión intra-grupos de \hat{y}_{it}^* en x_{it} . Transformando las variables en [8] en desviaciones con respecto a las medias, los η_i 's son eliminados. Así, se pueden tomar esperanzas condicionales

$$E(\hat{y}_{it}^* | x_{it}^T) = \beta \tilde{x}_{it}. \quad [26]$$

Según [24], $E(y_{it}^* | x_{it}^T) = \pi_i^* x_{it}^T$, por tanto, se tiene la siguiente expresión para las restricciones:

$$\pi_i^* x_{it} = \beta \tilde{x}_{it}, \quad [27]$$

donde $\pi_i^* = \pi_i - (\pi_1 + \dots + \pi_T)/T$. Así, β resuelve la siguiente expresión:

$$\beta = \left[\sum_i \sum_t \tilde{x}_{it} \tilde{x}_{it}' \right]^{-1} \sum_i \sum_t \tilde{x}_{it} \pi_i^* x_{it}, \quad [28]$$

por lo que se pueden obtener estimaciones consistentes de β simplemente reemplazando π_i^* en [28] por un estimador consistente $\hat{\pi}_i^*$.

Sin embargo, la desventaja de este método es la misma que la del modelo de Chamberlain: se necesita una especificación explícita de la distribución de η_i condicional en x_{it}^T . Se ha elegido una forma lineal, pero probablemente es una aproximación muy pobre. Además, esta clase de estimadores de distancia mínima también requieren la existencia de variables estrictamente exógenas. Esta es una importante limitación de estos estimadores, al igual que del de Honoré-Kyriazidou, ya que en muchos casos modelizar las relaciones dinámicas entre las variables constituye un aspecto esencial del modelo económico de interés⁵.

Arellano y Carrasco (1996) propusieron un modelo de efectos aleatorios semiparamétrico, en el que se permite que las variables explicativas sean predeterminadas y no estrictamente exógenas. Consideraron el siguiente modelo Probit para datos de panel:

$$y_{it} = 1 (\gamma y_{it-1} + \beta x_{it} + v_{it} \geq 0) \quad (i = 1, \dots, N; \quad t = 1, \dots, T, \quad [29]$$

$$v_{it} = \eta_i + u_{it},$$

donde $1(\cdot)$ representa la función indicador. Se supone que los errores compuestos v_{it} tienen una distribución normal de la forma:

⁵ Un ejemplo de esta situación viene dado por el efecto de los hijos en la participación laboral de la mujer (véase BROWNING, 1992).

$$v_{it} | w_i^t \sim N(E(\eta_i | w_i^t), \sigma_\epsilon^2), \quad [30]$$

donde $w_{it} = (x_{it}, y_{it-1})$ y $w_i^t = (w_{i1}, \dots, w_{it})$. Este modelo se puede considerar como un tipo de modelo de efectos aleatorios semiparamétrico, ya que contiene una esperanza condicional de los efectos individuales no paramétrica, siendo por lo demás un modelo paramétrico. La secuencia de medias condicionadas $\{E(\eta_i | w_i^s), s = 1, \dots, T\}$ se deja sin restringir, excepto por el hecho de que están vinculadas por la ley de las esperanzas iteradas. Así pues, el modelo permite que exista dependencia entre la variable explicativa x_{it} y el efecto individual η_i a través de la media condicional de esta última dada la senda observada de w . Además, el modelo especifica x como una variable predeterminada, en el sentido de que aunque x_{it} no depende de valores presentes o futuros del error u_{it} , puede haber dependencia entre valores retardados de u hacia x_{it} . El supuesto de normalidad no es esencial y puede ser reemplazado por cualquier otra especificación paramétrica. La probabilidad inicial $Pr(y_{i0} = 1)$ se deja sin restringir, por lo que se evita, también, el problema de las condiciones iniciales y únicamente se añade un parámetro a la función de verosimilitud.

El modelo puede estimarse por MV. La función de verosimilitud se maximiza como función de los parámetros sujeta a las restricciones en $E(\eta_i | w_i^t)$. También puede estimarse por MGM. Este método es más sencillo, ya que evita la estimación conjunta de los parámetros de interés y de los coeficientes $E(\eta_i | w_i^t)$. Arellano y Carrasco (1996) aplicaron este modelo a la estimación de una ecuación de participación laboral femenina teniendo en cuenta que las variables de hijos son predeterminadas y que existen efectos individuales. Sus resultados indican que el hecho de tratar los hijos como variables predeterminadas y no estrictamente exógenas da lugar a un efecto más negativo de éstos sobre la probabilidad de participación.

Como comentario final, es importante destacar que la decisión de tratar los efectos individuales como fijos o como aleatorios no es sencilla. Por una parte, los modelos de efectos fijos son atractivos, ya que no requieren la especificación de la distribución de los efectos inobservables, mientras que el enfoque de efectos aleatorios sí requiere especificar dicha distribución. Sin embargo, una desventaja del modelo de efectos fijos es que sólo puede utilizarse para ciertas distribuciones, y normalmente requiere supuestos más restrictivos sobre la distribución de las perturbaciones que tienen variación temporal. Por tanto, existe un equilibrio entre imponer restricciones en la distribución de los efectos individuales invariantes en el tiempo o en la distribución de las perturbaciones. Sin embargo, no debe olvidarse que la cuestión relevante es si los efectos están, o no, correlacionados con las variables explicativas. Si no lo están, el modelo puede identificarse utilizando únicamente una sección cruzada ($T = 1$) y, en ese caso, sería conveniente considerar el enfoque de los modelos de componentes del error.

2.3. Correlación serial o dependencia de los estados

Una cuestión importante que surge en el contexto de modelos lineales es la de distinguir entre correlación serial y dependencia de los estados. En este caso, la pregunta relevante es si existen efectos directos adicionales de la variable dependiente retardada aparte

de los generados indirectamente por la correlación serial de los errores. Para analizar esta cuestión, consideremos los dos modelos siguientes:

$$y_{it}^* = \gamma_{it-1} + \eta_i + e_{it}, \quad [31]$$

y

$$y_{it}^* = \eta_i + u_{it}, \quad u_{it} = \rho u_{it-1} + e_{it}, \quad [32]$$

donde, en ambos casos, $y_{it} = 1$ si $y_{it}^* \geq 0$, $y_{it} = 0$ en caso contrario, y e_{it} es $iid N(0, \sigma^2)$.

El modelo [31] es un modelo en el que hay dependencia de los estados, mientras que en el modelo [32] hay correlación serial. En el modelo [31]

$$Pr(y_{it} = 1 | \eta_i, y_{it-1}, y_{it-2}, \dots) = Pr(y_{it} = 1 | \eta_i, y_{it-1}) = F(\gamma_{it-1} + \eta_i), \quad [33]$$

y, por tanto, implica una cadena de Markov de primer orden. Sin embargo, en el modelo [32], esta probabilidad depende del historial completo del proceso.

Es importante destacar la importancia de tener en cuenta la heterogeneidad inobservable para distinguir la correlación serial de la dependencia de los estados. Dado que los efectos individuales inobservables tienden a persistir en el tiempo, ignorarlos generará residuos serialmente correlacionados.

Si una vez que se condiciona en los efectos individuales η_i , el término de error e_{it} está serialmente incorrelacionado, es fácil implementar un test para contrastar si existe o no dependencia de los estados, contrastando si la probabilidad condicional es igual a la marginal:

$$Pr(y_{it} | y_{it-1}, \eta_i) = Pr(y_{it} | \eta_i). \quad [34]$$

Cuando, por el contrario, una vez se condiciona a los efectos individuales, el término de error permanece serialmente correlacionado, el problema es más complicado. En este caso, la distinción entre los modelos [31] y [32] se basa en el orden de la dependencia de las realizaciones anteriores de y . No obstante, Chamberlain (1978) considera que basar el test en el orden de la autoregresión no es muy adecuado. Si existen datos disponibles sobre x_t , podemos reformular el test preguntándonos si existe respuesta dinámica ante cambios en x_t o no. Consideremos

$$y_{it}^* = \gamma_{it-1} + \beta x_{it} + \eta_i + e_{it}. \quad [35]$$

Un test de dependencia de los estados se realizaría incluyendo valores retardados de x 's y contrastando si sus coeficientes son significativamente distintos de 0 o no. Si $\gamma = 0$, un cambio en x produce un efecto completo inmediato, mientras que $\gamma \neq 0$ implicaría una respuesta distribuida ante un cambio en x . Una vez se condiciona en el efecto individual específico η_i , pueden darse dos resultados. Si no hay dependencia de los estados, entonces

$$Pr(y_{it} | x_{it}, x_{it-1}, \dots, \eta_i) = Pr(y_{it} | x_{it}, \eta_i), \quad [36]$$

mientras que si hay dependencia de los estados, entonces

$$Pr(y_{it} | x_{it}, x_{it-1}, \dots, \eta_i) \neq Pr(y_{it} | x_{it}, \eta_i). \quad [37]$$

Un problema que surge con la correlación serial es que depende crucialmente del intervalo muestral. Cuanto menor es dicho intervalo, mayor es la correlación serial. Chamberlain (1984) afirma que el hecho de que el coeficiente γ sea significativo puede no decir mucho sobre el proceso subyacente. De hecho, en muchos casos, sería más conveniente utilizar modelos de duración, que a continuación se analizan.

3. Modelos de duración

3.1. Datos de duraciones

Los datos de duraciones miden cuánto tiempo permanece un individuo en un estado determinado. Estos datos registran la secuencia de estados que han sido ocupados por el individuo, así como las transiciones registradas entre estados. Algunos ejemplos que se suelen considerar son las secuencias de empleo, desempleo e inactividad de un individuo, o las secuencias de empleo y jubilación, o de matrimonio y divorcio (véanse Kalbfleisch y Prentice, 1980; Lancaster, 1990, para una descripción detallada de los modelos de duración).

Hay varias razones por las que los datos de duración son peculiares. Una de sus características es que existe la posibilidad de que algunas de las duraciones observadas estén censuradas. Por ejemplo, en muchas ocasiones no se observa la finalización de los periodos de desempleo de los individuos. Se dice que las observaciones de este tipo están «censuradas por la derecha». En otras ocasiones, la duración real tampoco se conoce, debido a que no se conoce el tiempo que transcurre entre el momento de inicio del periodo de desempleo y el momento en el que se empiezan a recoger los datos. En este caso, la duración estaría «censurada por la izquierda». Otro problema es que, en algunas ocasiones, la información sobre el periodo de desempleo sólo se recoge para aquellos individuos que ya están desempleados en el momento de la entrevista. Esto significa que no se tienen datos para los individuos que están empleados en el momento de realizarse la entrevista, pero que están desempleados entre dos entrevistas. Este problema, cuya consecuencia es que las duraciones cortas puedan estar subrepresentadas en la muestra, se conoce como «sesgo de stock».

Dado que es necesario tener en cuenta estos problemas de censura, al considerar estos modelos se necesita disponer de información sobre la duración de un determinado suceso, así como sobre si la duración está censurada o no. Sea T^* , una variable aleatoria que representa una determinada duración sin censura, y sea C el momento en el que se produce la censura. Entonces, la variable aleatoria observada será igual a $T = \min(T^*, C)$. También será necesario disponer de una variable indicador $c = 1$ si la observación es censurada ($T = C$), $c = 0$ si es no censurada ($T = T^*$).

No obstante, incluso si la muestra consistiera exclusivamente en una muestra aleatoria de duraciones completas, es decir, incluso si no hubiera censura en los datos, utilizar la información sobre duraciones de desempleo como si se tratara de una variable dependiente en un contexto de regresión donde los determinantes de la duración se miden por un con-

junto de variable exógenas x presenta dificultades. El problema en este contexto sería cómo medir aquellas x 's cuyos valores cambian durante el periodo de desempleo. Por lo tanto, incluso sin tener en cuenta el problema de la censura, los datos de duración presentan problemas conceptuales si estamos acostumbrados a pensar en términos del análisis de regresión convencional.

Por estas razones, ha surgido una literatura que trata los problemas específicos asociados a los datos de duración. El análisis de estos datos tiene una larga tradición en las ciencias biomédicas, siendo utilizados para describir hechos tales como el tiempo de supervivencia de pacientes con trasplantes de corazón. Sin embargo, estos métodos también tienen una aplicación natural a muchos problemas económicos. Probablemente, los datos de duración más ampliamente estudiados en economía son los de las duraciones de desempleo (véanse Lancaster, 1979; Nickell, 1979; Kiefer y Neumann, 1979; Narendranathan y Nickell, 1985; Bover, Arellano y Bentolila, 1996, entre otros).

3.2. Modelos de duración en tiempo discreto

Los datos económicos convencionales, normalmente, proporcionan información sobre duraciones que están agregadas en intervalos discretos. Por ejemplo, se suelen observar duraciones de desempleo en semanas o trimestres, en lugar de como realizaciones continuas de la variable T . Algunos autores opinan que los modelos en tiempo continuo son más apropiados en el contexto de los problemas económicos, ya que no es posible determinar un periodo natural en el que se tomen las decisiones económicas. Además, con los modelos en tiempo continuo, los parámetros pueden interpretarse independientemente del periodo al que se refieren los datos. Sin embargo, en algunos casos, el enfoque de tiempo discreto es más adecuado. Los datos de duración son a veces discretos, bien porque el tiempo es en sí mismo discreto o porque los datos continuos se agrupan debido a la existencia de problemas de medición⁶. Por otra parte, para el análisis empírico de datos de duración, los modelos en tiempo discreto presentan algunas ventajas. Una de ellas es que se pueden estimar de forma muy sencilla, teniendo en cuenta tanto variables que varían en el tiempo, como especificaciones muy flexibles de la dependencia de la duración. Además, este marco pone en evidencia la estrecha correspondencia existente entre modelos de duración y modelos binarios de elección discreta para datos de panel. Por todo ello, los modelos en tiempo discreto son ampliamente utilizados en economía.

Consideremos, entonces, el caso en el que T es una variable aleatoria discreta que toma los valores $\{1, 2, 3, \dots\}$. La función de riesgo en el momento t se define como la probabilidad condicional de abandonar un estado determinado, dado que no se ha abandonado hasta el momento t :

$$h(t) = \Pr(T = t \mid T \geq t). \quad [38]$$

La función de supervivencia asociada a esta función de riesgo viene dada por

$$S(t) = 1 - F(t) = \Pr(T > t). \quad [39]$$

⁶ Los datos de duración discreta de este tipo se conocen como «datos agrupados» (KIEFER, 1988).

Entonces,

$$p(t) = \Pr(T = t), \quad [40]$$

y la función de distribución acumulada es:

$$F(t) = \Pr(T \leq t) = p(1) + p(2) + \dots + p(t). \quad [41]$$

El concepto fundamental en estos modelos viene dado por la probabilidad condicional de que un suceso se produzca. Esto es, la función de riesgo pone el énfasis en las probabilidades condicionales, mientras que la especificación en términos de la distribución de probabilidad enfatiza las probabilidades incondicionales. Por ejemplo, si T representa la duración del desempleo, $h(t)$ es la probabilidad de abandonar el desempleo en t dado que el individuo ha estado desempleado al menos t periodos. Las dos especificaciones implican los mismos parámetros y son simplemente dos formas diferentes de describir el mismo sistema de probabilidades. Veamos cómo, en efecto, es posible recuperar $F(t)$ y $f(t)$ a partir de $h(t)$. Nótese que para $t > 1$

$$1 - h(t) = \frac{1 - F(t)}{1 - F(t-1)}. \quad [42]$$

Por tanto,

$$1 - F(t) = (1 - F(t-1)) (1 - h(t)),$$

y

$$1 - F(t) = \prod_{s=1}^t (1 - h(s)), \quad t = 1, 2, \dots \quad [43]$$

De forma similar,

$$p(t) = (1 - F(t-1)) h(t). \quad [44]$$

Dada esta equivalencia, las contribuciones a la función de verosimilitud para una muestra de individuos con variables de duración discreta pueden basarse en la especificación de la función de riesgo h .

3.2.1. Estimación de modelos de duración discretos como una secuencia de modelos binarios

Los modelos de duración en tiempo discreto pueden considerarse como una secuencia de ecuaciones de elección discreta definidas sobre la población superviviente en cada dura-

ción⁷. Por tanto, estos modelos pueden estimarse como modelos de elección binaria, utilizando los programas estándar disponibles para ello.

Sea T_i la variable de duración censurada observada, de forma que:

$$T_i = \begin{cases} T_i^* & \text{si } T_i^* < C_i, \\ C_i & \text{si } T_i^* > C_i, \end{cases} \quad [45]$$

donde C_i es el número de periodos que el individuo permanece en la muestra. Sea c_i el indicador de ausencia de censura

$$c_i = 1 (T_i^* < C_i). \quad [46]$$

Definamos una variable (0, 1) indicando si la duración observada es igual a t o no:

$$y_{it} = 1 (T_i = t), \quad t = 2, 3, \dots \quad [47]$$

y

$$w_{it} = 1 (T_i \geq t). \quad [48]$$

Así, para individuos que no cambian de estado $y_{it} = 0$ en todos los periodos; para los individuos que cambian de estado $y_{it} = 0$ para todos los periodos, excepto para el periodo en el que se produce el cambio. Utilizando estas variables indicador, la función de riesgo puede escribirse como

$$h_i(t) = Pr(y_{it} = 1 | T_i \geq t) = Pr(y_{it} = 1 | w_{it} = 1). \quad [49]$$

Dado que normalmente estaremos preocupados por analizar la relación entre la duración y un conjunto de variables explicativas x_i , se puede utilizar la función de riesgo condicional

$$h_i(t, x_i(t)) = Pr(y_{it} = 1 | w_{it} = 1, x_i(t)) = F(\gamma_t + x'_{it}\beta), \quad [50]$$

donde $F(\cdot)$ es una fda. Por tanto, se permite que la función de riesgo varíe entre individuos a través de una parametrización que depende de variables observables x y de parámetros β .

La función de log-verosimilitud de la muestra para y_{it} dado $T_i \geq t$ toma la forma:

$$L_t = \sum_{i=1}^N w_{it} \{c_i y_{it} \log F(\gamma_t + x'_{it}\beta) + (1 - c_i y_{it}) \log [1 - F(\gamma_t + x'_{it}\beta)]\}. \quad [51]$$

⁷ Véanse KIEFER (1987), NARENDRANATHAN y STEWART (1993), SUEYOSHI (1995) y JENKINS (1995).

donde N es el número de duraciones en la muestra. Combinando las L_t para todas las duraciones observadas, se obtiene el siguiente criterio de estimación:

$$L = \sum_{t=1}^{t^*} L_t = \sum_{i=1}^N \left\{ (1 - c_i) \sum_{t=1}^{T_i} \log [1 - F(t)] + c_i \left(\sum_{t=1}^{T_i-1} \log [1 - F(t)] + \log [F(T_i)] \right) \right\}, \quad [52]$$

donde t^* es la mayor duración observada. Por tanto, en este modelo las salidas o permanencias en cada periodo se consideran como una observación. Así, la función de verosimilitud tiene la misma forma que la función de verosimilitud estándar del análisis de variables binarias, en este caso y_{it} .

En el caso más general, γ_t y β_t se dejarían sin restringir y este modelo sería equivalente a una serie de funciones de verosimilitud de t^* modelos de elección binaria definidos para la población superviviente en cada t . Para completar la especificación de la función de verosimilitud, se ha de definir una forma funcional para la función de riesgo. Las especificaciones más frecuentemente utilizadas para F son la logística, la normal y también la distribución de valor extremo:

$$h_i(t) = 1 - \exp [- \exp (\gamma_t + x_{it}'\beta)]. \quad [53]$$

Esta última tiene la propiedad de que el modelo resultante es la contrapartida en tiempo discreto de un modelo continuo subyacente de riesgo proporcional (véase Prentice y Gloeckler, 1978). En el modelo de riesgo proporcional (Cox, 1972) las diferencias absolutas en las variables implican diferencias proporcionales entre las tasas de riesgo. Este modelo especifica que:

$$\theta_i(t) = \lambda(t) \exp (x_i'(t) \beta), \quad [54]$$

donde $\lambda(t)$ es el «riesgo base» en t . En esta especificación el efecto de las variables explicativas consiste en multiplicar el riesgo λ por un factor que no depende de la duración t . Por tanto, la función de riesgo se factoriza en una función de t y en una función de x , de forma que dos individuos diferentes tendrían, por ejemplo, probabilidades de reempleo proporcionales para todo t .

Parte de la literatura existente sobre datos de duración ha enfatizado la estrecha relación que hay entre el estimador de máxima verosimilitud para datos agrupados y las especificaciones de modelos binarios tradicionales (Kiefer, 1988; Sueyoshi, 1991). Según esta literatura, es posible trasladar las especificaciones de duraciones en tiempo continuo a un contexto de datos de duración agrupados. La verosimilitud de una observación particular de un modelo de duración agrupado es, simplemente, la probabilidad de observar una serie de resultados binarios, con probabilidades dadas por la función de distribución de valor extremo. Para entender esta idea, consideremos la variable de duración (T) como una variable aleatoria continua que se observa a intervalos discretos.

Supongamos que la duración exacta es desconocida, y que lo único que se observa es el intervalo durante el que se produce la duración. Así, una duración observada de t periodos indica una duración en tiempo discreto de entre t y $t+1$ periodos. Se puede demostrar que, en tiempo continuo, la probabilidad de que una duración se haya completado en el periodo $t+1$ dado que aún no ha terminado en el periodo t , viene dada por la siguiente expresión:

$$\begin{aligned} h_i(t) &= \Pr [T_i < t+1 \mid T_i \geq t] = 1 - \exp \left[- \int_t^{t+1} \theta_i(u) du \right] \\ &= 1 - \exp \left[- \int_t^{t+1} \lambda(u) \exp(x_i(u)' \beta) du \right]. \end{aligned} \quad [55]$$

Suponiendo que $x_i(u)$ es constante para $t \leq u < t+1$, la función de riesgo en tiempo discreto puede escribirse como un modelo de valor extremo:

$$h_i(t) = 1 - \exp \left[- \exp \{ \gamma(t) + x_i(u)' \beta \} \right], \quad [56]$$

donde

$$\gamma(t) = \ln \left\{ \int_t^{t+1} \lambda(u) du \right\}$$

es un parámetro no restringido específico para cada t y que recoge una dependencia de la duración aditiva. Así pues, el modelo en tiempo discreto toma la forma de una distribución de valor extremo⁸. Una de las ventajas de elegir esta especificación es que las estimaciones son parámetros de un modelo en tiempo continuo, lo que facilita su interpretación. Otra ventaja es que de esta forma es posible estimar el riesgo base no paramétricamente, mientras que el enfoque convencional requiere utilizar una especificación paramétrica (véase Meyer, 1990).

3.3. La heterogeneidad inobservable

En los modelos anteriores se ha supuesto que toda la heterogeneidad individual se debe a variables observables. Sin embargo, es probable que las variables inobservables constituyan también un factor de heterogeneidad entre los individuos. La consecuencia de este hecho en modelos de duración es producir inferencias erróneas sobre la dependencia de la duración y sobre el resto de las variables explicativas (véanse Davies y Pickles, 1985; y Ridder, 1987). En concreto, el hecho de no controlar por la heterogeneidad inobservable produce un sesgo negativo en la dependencia de la duración.

El procedimiento más extendido para tener en cuenta la heterogeneidad inobservable consiste en asumir que se puede representar mediante la introducción en la función de riesgo de un término de perturbación estocástico, v , con función de densidad $g(v)$. Así, la función de riesgo se especifica como

$$h^*(t, x, v) = h(t, x) v. \quad [57]$$

⁸ Véanse MEYER (1990) y NARENDRANATHAN y STEWART (1993).

Un enfoque natural para acomodar la existencia de heterogeneidad inobservable consiste en basar la inferencia en la distribución mixta que resulta de la presencia de la heterogeneidad. Fundamentalmente hay dos enfoques para modelizar esta distribución mixta. El primero consiste en asumir que v_i sigue alguna distribución paramétrica e integrarla fuera de la verosimilitud (por ejemplo Lancaster, 1979, supone una distribución gamma). El segundo enfoque consiste en aproximar la distribución continua por una distribución discreta de forma no restringida (Heckman y Singer, 1984).

Es importante destacar que para datos de un único periodo de duración de un proceso no es posible tener en cuenta la heterogeneidad inobservable sin especificar una distribución. Sin embargo, la existencia de datos de panel de duraciones permitiría considerar algunos enfoques alternativos para controlar por las variables inobservables, sin necesidad de especificar una distribución para los efectos individuales inobservables. En ese caso, sería posible tenerlos en cuenta de forma similar a como se hace en los modelos de elección discreta para datos de panel (por ejemplo, basando la inferencia en la verosimilitud condicional).

3.4. Algunas generalizaciones: riesgos en competencia y duraciones múltiples

Los datos de duraciones pueden ser más generales que los presentados en las secciones anteriores. En primer lugar, puede existir más de una causa de abandono de un estado determinado. Son los llamados modelos de riesgos en competencia. En segundo lugar, puede observarse más de una duración para cada individuo. Son los llamados modelos de duraciones múltiples.

3.4.1. Modelos de riesgos en competencia

Los modelos presentados previamente especifican los determinantes de la salida de un estado, por ejemplo, salida del desempleo, sin especificar el destino. Consideremos ahora una situación en la que hay riesgos en competencia, es decir, en la que, además de las duraciones de desempleo, se especifica el estado al que el individuo sale o la causa del abandono (entrada al empleo, al autoempleo, etcétera)⁹.

La estimación de este tipo de modelos es muy sencilla, ya que la función de log-verosimilitud es aditivamente separable en términos que son función, únicamente, de los parámetros de una sólo causa de salida¹⁰. Denotemos la j -ésima ($j = 1, 2, \dots, J$) causa de salida del individuo i por $h_{ji}(\cdot)$. La función de riesgo vendrá dada por:

$$h_i(t) = \sum_{j=1}^J h_{ji}(t). \quad [58]$$

⁹ Véanse KATZ y MEYER (1988), NARENDRANATHAN y STEWART (1993) o CARRASCO (1997).

¹⁰ Esta aditividad separable se pierde en modelos que contienen términos de heterogeneidad omitidos correlacionados entre los distintos riesgos.

Sean las variables indicador $c_{ji} = 1$ si i sale al estado j ; 0 en otro caso ($j = 1, \dots, J$). Nótese que $c_i = \sum_j c_{ji}$. Por tanto, las contribuciones a la función de log-verosimilitud vienen dadas por:

$$L_j = \sum_{i=1}^N \left[c_{ij} \left(\sum_{t=1}^{d_i-1} \{ \log [1 - h_{ij}(t)] + \log h_{ij}(d_i) \} \right) + (1 - c_{ij}) \left(\sum_{t=1}^{d_i} \log [1 - h_{ij}(t)] \right) \right], [59]$$

donde d_i es la duración observada para el individuo i (completada o censurada). La función de verosimilitud será la suma de términos como [59] para todo j .

Así pues, si los diferentes destinos dependen de distintos subconjuntos de parámetros independientes, entonces, por lo que a la inferencia sobre $(\gamma_j(t), \beta_j)$ se refiere, la verosimilitud puede tomarse simplemente como L_j dada por [59]. Por tanto, es posible particionarla en una suma de términos, cada uno de los cuales es función de los parámetros de una única causa de salida. Así, los parámetros de cada causa de salida pueden estimarse tratando las duraciones que terminan por alguna otra causa como censuradas en ese momento. Sin embargo, a pesar de la sencillez de este método de estimación, el inconveniente es que, si estamos interesados en contrastar hipótesis que implican restricciones entre los riesgos, es necesario estimarlos conjuntamente. En este sentido, una restricción interesante es la impuesta por el modelo de riesgo proporcional: la igualdad de los coeficientes del riesgo base, hasta un factor de escala. Narendranathan y Stewart (1991) presentaron un método sencillo para contrastar esta hipótesis.

3.4.2. Modelos de duraciones múltiples: heterogeneidad en modelos de duración

En los datos de duración discutidos hasta ahora se observa una única duración para cada individuo. Modelos de comportamiento más ricos son aquéllos en los que se observan distintas duraciones para el mismo individuo. Se podría considerar también una situación general en la que existan riesgos en competencia y, además, múltiples duraciones.

Si se ignora la heterogeneidad inobservable, se puede demostrar que la función de verosimilitud de estos modelos puede factorizarse en distintos componentes para cada riesgo y para cada duración (véase Kalbfleisch y Prentice, 1980). Como en el caso anterior, se puede definir la función de riesgo específica para cada duración. La función de verosimilitud para una duración en concreto es, precisamente, la misma que la que se obtendría a partir de una muestra con una única duración y un único riesgo, estimándose los parámetros de la misma forma que en las secciones anteriores.

El modelo de duraciones múltiples puede extenderse para tener en cuenta la heterogeneidad inobservable, η_i . Para ello, se puede escribir la función de verosimilitud marginal, integrando los efectos individuales (véase Flinn y Heckman, 1982). Sin embargo, y a diferencia de lo que ocurre con datos de una única duración, también es posible escribir la verosimilitud condicional, con la ventaja de que no es necesario realizar supuestos paramétricos sobre la distribución de la heterogeneidad.

Chamberlain (1984) presentó una clase de modelos de duración múltiple para los que es posible encontrar estadísticos suficientes para η_i . La desventaja de su procedimiento

es que no proporciona estimaciones de los parámetros de la dependencia de la duración y requiere supuestos restrictivos sobre la forma funcional del riesgo condicional. Por tanto, el desarrollo de condiciones bajo las que es posible separar el efecto de la dependencia de la duración del efecto de la heterogeneidad inobservable con un procedimiento no paramétrico constituye un campo de investigación futura. Chamberlain también consideró algunos casos especiales en los que es posible utilizar un enfoque de efectos fijos. Sin embargo, en estos modelos se asume la independencia de las distintas duraciones. Otra importante línea de investigación consiste, por tanto, en desarrollar modelos que relajen este supuesto.

4. Secuencias binarias y secuencias de estados

Los modelos de elección discreta para datos de panel son apropiados cuando el periodo de observación tiene algún significado natural y cuando el intervalo temporal se corresponde con decisiones de los individuos. Situaciones en las que un individuo votase a un partido de una ideología u otra, o en las que una empresa repartiese dividendos, o no, en cada periodo, generarían una secuencia de decisiones binarias que proporcionarían una descripción completa del proceso de interés. En muchas ocasiones, sin embargo, los datos consisten, únicamente, en una secuencia de estados que se producen a intervalos discretos, junto con el tiempo que el individuo permanece en cada estado. Por ejemplo, en el estudio de la participación en el mercado de trabajo, una descripción completa del proceso vendría dada por la duración del primer periodo de empleo junto con su fecha de inicio, seguido de la duración del siguiente periodo de desempleo, y así sucesivamente. Esta información generará una secuencia binaria cuando se corta en intervalos fijos de tiempo, pero estos intervalos no nos dicen mucho sobre el proceso subyacente.

En particular, en modelos de elección binaria que incluyen la variable dependiente retardada como un regresor adicional, la medida de correlación serial depende del intervalo muestral: cuanto menor sea éste, mayor será la correlación serial. Por ejemplo, supongamos que nuestro periodo de observación es un día. La probabilidad de que una persona que estaba empleada ayer siga empleada hoy estará muy cercana a 1. Chamberlain (1984) afirma que el hecho de encontrar evidencia de correlación serial en un caso como éste nos diría muy poco sobre el proceso subyacente. Por tanto, una cuestión más interesante sería analizar si el historial del individuo ayuda a predecir su situación futura dado su estado actual, por lo que el interés se centraría en estudiar los factores que afectan a las transiciones entre estados.

En este contexto, sería más apropiado utilizar modelos de duración en tiempo discreto en lugar de modelos de elección discreta. Las funciones de verosimilitud para ambos modelos son muy similares, sin embargo, el contenido en términos del comportamiento de los individuos que implica cada modelo es muy diferente. Incluso aunque en ambos casos los datos consisten en secuencias binarias, la interpretación es muy diferente. En un caso representan secuencias de decisiones individuales y en el otro representan secuencias de estados. En el primer caso podemos usar modelos de elección discreta para datos de panel, mientras que en el segundo es más apropiado utilizar modelos de duración.

Supongamos que la variable y_{it} indica el estado ocupado por un individuo, empleo o desempleo, en el momento t :

$$y_{it} = \begin{cases} 1 & \text{si } i \text{ está empleado en } t, \\ 0 & \text{si } i \text{ está desempleado en } t. \end{cases} \quad [60]$$

Podemos especificar el siguiente modelo para y_{it} :

$$Pr(y_{it} = 1 \mid y_{i1}, \dots, y_{it-1}) = F(\gamma + \alpha y_{it-1}) = F_{it}, \quad [61]$$

donde $F(\cdot)$ es una fda. Por tanto, las probabilidades especificadas por el modelo son las siguientes:

$$F_{it} = \begin{cases} F(\gamma + \alpha) = P_{11} & \text{si } y_{it-1} = 1, \\ F(\gamma) = P_{01} & \text{si } y_{it-1} = 0. \end{cases} \quad [62]$$

Consideremos un trabajador empleado en el periodo $t - 1$. La probabilidad de que continúe empleado en el periodo siguiente viene dada por P_{11} . Para un trabajador desempleado en $t - 1$ la probabilidad de que esté empleado en t es $1 - P_{01}$. Como se ve en el modelo, las transiciones dependen solamente del estado actual. Por tanto, el modelo autoregresivo de primer orden implicaría tasas de salida constantes: no existe dependencia de la duración, y el tiempo que el individuo permanece en un estado no afecta a la tasa de salida de dicho estado. Para poder tener en cuenta la dependencia de la duración se deberían incluir retardos adicionales de y como regresores.

Dado que la correlación serial depende del intervalo muestral entre observaciones, Chamberlain argumenta que un punto de partida más interesante consistiría en analizar si la historia pasada del individuo ofrece información sobre su estado actual. Si éste fuera el caso, estaríamos hablando de un proceso de Markov. En otro caso, habría dependencia de la duración.

Según Chamberlain (1984), es posible contrastar la dependencia de la duración utilizando secuencias de datos binarios generadas a partir de diferentes tipos de información muestral. El caso más sencillo es aquel en el que se observa el estado que ocupa cada individuo en cada momento del tiempo. Otro caso sería aquel en el que se le pregunta a cada individuo sobre su situación en un intervalo de tiempo pasado (por ejemplo, el año anterior).

El test de Chamberlain está basado en los modelos Logit condicionales. Para el primer tipo de datos, sugiere el siguiente modelo:

$$Pr(y_{it} = 1 \mid y_{it-1}, y_{it-2}) = \frac{\exp(\alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \eta_i)}{1 + \exp(\alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \eta_i)}. \quad [63]$$

Los parámetros η_i y α_{i1} permiten que cada individuo tenga cadenas de Markov de primer orden específicas. Por tanto, la importancia y significatividad de la dependencia de la duración puede medirse mediante la estimación de α_2 y su intervalo de confianza. Chamberlain

obtiene estimaciones para α_2 utilizando el enfoque de la verosimilitud condicional, siendo $y_1, y_2, \sum_{i=1}^T y_i, \sum_{i=2}^T y_i y_{i-1}, y_{T-1}, e y_T$ los estadísticos suficientes para α_2 .

Sin embargo, aunque el modelo [61] permite la posibilidad de que un individuo desempleado tenga una probabilidad de seguir estándolo en el periodo siguiente distinta a la de un individuo empleado, no es muy atractivo modelizar conjuntamente ambas tasas de salida, $1 - P_{11}$ y P_{01} . A pesar de que ambas presentan un paralelismo estadístico (nótese que sin condicionar en y_{t-1} coinciden) el proceso subyacente que caracteriza el comportamiento de los distintos trabajadores es también muy diferente. El primero es un proceso de desempleo que opera cuando el individuo está desempleado, mientras que el segundo es un proceso de empleo. En un intervalo corto de tiempo, la probabilidad de que un trabajador empleado entre en el desempleo será diferente de la probabilidad de que un desempleado permanezca en este estado. Por tanto, sería deseable estimar separadamente P_{11} y P_{01} . Este problema se puede solucionar si se dispone de datos de duración. Chamberlain (1984) discute modelos para datos de panel de duraciones. Se centra en el estudio de la presencia de heterogeneidad inobservable en estos modelos. Demuestra que es posible utilizar tanto las técnicas de efectos fijos, como de efectos aleatorios, de forma similar al caso de modelos de elección discreta para datos de panel.

5. Conclusiones

En este artículo se han revisado algunos de los problemas que surgen en la estimación de modelos de elección discreta para datos de panel. El énfasis se ha puesto en presentar métodos que permiten eliminar el sesgo debido a la presencia de heterogeneidad inobservable. Se han analizado cuestiones relacionadas con los modelos de efectos fijos y de efectos aleatorios. La ventaja de las técnicas de efectos fijos es que no requieren supuestos paramétricos específicos sobre la distribución condicional de la heterogeneidad dadas las variables explicativas. La desventaja es que estas técnicas sólo pueden usarse para ciertas distribuciones. Por el contrario, el enfoque de efectos aleatorios puede utilizarse siempre, aunque a costa de imponer distribuciones paramétricas a veces muy restrictivas. Además, la cuestión de cómo la distribución de la heterogeneidad debe depender de las variables explicativas es aún objeto de investigación futura.

También se ha considerado el problema de las condiciones iniciales en modelos de elección discreta dinámicos, así como la cuestión de la correlación serial frente a la dependencia de los estados. Dado que la correlación serial depende del intervalo muestral, en algunas ocasiones es más adecuado analizar la dependencia de la duración. Esta cuestión puede ser estudiada directamente si se dispone de datos sobre la duración de los estados que ocupa un individuo. Por ello, se han considerado los modelos de duración en tiempo discreto y cómo éstos pueden estimarse como una secuencia de modelos de elección binaria, enfatizando la relación entre ambos tipos de modelos. También se han discutido los modelos para datos de panel de duraciones y cómo pueden ser estimados teniendo en cuenta la heterogeneidad inobservable entre individuos, por ejemplo, mediante el enfoque de efectos fijos. No obstante, ésta es todavía un área de investigación abierta.

Finalmente, se han descrito situaciones en las que es más adecuado utilizar modelos de duración, en lugar de modelos de elección discreta, dependiendo de la naturaleza de los datos disponibles.

Referencias bibliográficas

- [1] ANDERSEN, E. B. (1973): *Conditional Inference and Models for Measuring*, Mentalhygiejnisk Forsknings Institut, Copenhagen.
- [2] ANDERSON, T. W., y HSIAO, C. (1981): «Estimation of Dynamic Models with Error Components», *Journal of the American Statistical Association*, 76, 598-606.
- [3] ARELLANO, M., y BOND, S. (1991): «Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations», *Review of Economics Studies*, 58, 277-97.
- [4] ARELLANO, M., y BOVER, O. (1995): «Another Look at the Instrumental Variable Estimation of Error-Components Models», *Journal of Econometrics*, 68, 29-51.
- [5] ARELLANO, M., y CARRASCO, R. (1996): «Binary Choice Panel Data Models with Predetermined Variables», CEMFI, *Working Paper*, 9618.
- [6] AVERY, R. B., HANSEN, L. P., y HOTZ, V. J. (1983): «Multiperiod Probit Models and Orthogonality Condition Estimation», *International Economic Review*, 24, 21-35.
- [7] BALESTRA, P., y NERLOVE, M. (1966): «Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demy for Natural Gas», *Econometrica*, 51(6), 1635-59.
- [8] BATES, C., y WHITE, H. (1987): «Efficient Instrumental Variable Estimation of Systems of Equations with Nonspherical Errors», *UCSD Discussion Paper*, #87-14.
- [9] BOVER, O., y ARELLANO, M. (1988): «Estimating Dynamic Limited Dependent Variable Models from Panel Data with an Application to Female Labour Supply», Nuffield College, Oxford.
- [10] BOVER, O., ARELLANO, M., y BENTOLILA, S. (1996): «Unemployment Duration, Benefit Duration and the Business Cycle», de próxima aparición en *The Economic Journal*.
- [11] BROWNING, M. (1992): «Children and Household Economic Behaviour», *Journal of Economic Literature*, 30, 1434-75.
- [12] BUTLER, J. S., y MOFFITT, R. (1982): «A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model», *Econometrica*, 50, 761-64.
- [13] CARD, D., y SULLIVAN, D. (1988): «Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment», *Econometrica*, 56, 497-530.
- [14] CARRASCO, R. (1997): «Transitions to and from Self-Employment in Spain: An Empirical Analysis», de próxima aparición en *Oxford Bulletin of Economics and Statistics*.

- [15] CHAMBERLAIN, G. (1980): «Analysis of Covariance with Qualitative Data», *Review of Economic Studies*, 47, 225-38.
- [16] CHAMBERLAIN, G. (1984): «Heterogeneity, Omitted Variable Bias and Duration Dependence», en J. HECKMAN y B. SINGER (eds.), *Longitudinal Analyses of Labor Market Data*, Nueva York: Academic Press.
- [17] CHAMBERLAIN, G. (1985): «Panel Data», en Z. GRILICHES y M. D. INTRILIGATOR (eds.), *Handbook of Econometrics*, vol. 2, 1258-318. Amsterdam: North-Holly Publishing Co.
- [18] COX, D. R. (1970): *Analysis of Binary Data*, Londres: Methuen.
- [19] COX, D. R. (1972): «Regression Models and Life-Tables», J. ROY, *Statist. Soc.*, 34, 187-220.
- [20] FLINN, C., y HECKMAN, J. (1982): «Models for the Analysis of Labor Force Dynamics», en R. BASMANN y G. RHODES (eds.), *Advances in Econometrics*, 1, 35-95.
- [21] HAJIVASSILIOU, V., y McFADDEN, D. (1990): «The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crisis», Coules Foundation, *Discussion Paper*, 967.
- [22] HAM, J. C., y REA, S. A. (1987): «Unemployment Insurance and Male Unemployment Duration in Canada», *Journal of Labor Economics*, 5, 325-53.
- [23] HECKMAN, J. J. (1981): «Statistical Models for Discrete Panel Data», en C. F. MANSKI y D. McFADDEN (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Mass.
- [24] HECKMAN, J. J., y SINGER, B. (1984): «A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data», *Econometrica*, 52(2), 271-320.
- [25] HECKMAN, J. J., y WILLIS, R. (1977): «A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women», *Journal of Political Economy*, 85(1), 27-58.
- [26] HOLTZ-EAKIN, D.; NEWE, W., y ROSEN, H. S. (1988): «Estimating Vector Autoregressions with Panel Data», *Econometrica*, 56, 1371-95.
- [27] HONORÉ, B., y KYRIAZIDOU, E. (1996): «Panel Data Discrete Choice Models with Lagged Dependent Variables», próxima aparición en *Econometrica*.
- [28] HOROWITZ, J. (1992): «A Smoothed Maximum Score Estimator for the Binary Response Model», *Econometrica*, 60, 505-31.
- [29] HSIAO, C. (1986): *Analysis of Panel Data*, Cambridge University Press.
- [30] JENKINS, S. (1995): «Easy Estimation Methods for Discrete Time Duration Models», *Oxford Bulletin of Economics and Statistics*, 57.1, 120-138.
- [31] KALBFLEICH, J., y PRENTICE, R. (1980): *The Statistical Analysis of Failure Time Data*, Nueva York: Wiley.
- [32] KATZ, L. F., y MEYER, B. D. (1988): «Unemployment Insurance, Recall Expectations and Unemployment Outcomes», NBER, *Working Paper*, 2594.
- [33] KEANE, M. P. (1994): «A Computationally Practical Simulation for Panel Data», *Econometrica*, 62(1), 95-116.

- [34] KIEFER, N. (1988): «Analysis of Grouped Duration Data», Cornell CAE, *Working Paper*, 87-12.
- [35] KIEFER, N., y NEUMANN, G. (1979): «An Empirical Job Search Model, with a Test of the Constant-Reservation Wage Hypothesis», *Journal of Political Economy*, 87(1), 89-107.
- [36] KYRIAZIDOU, E. (1997): «Estimation of a Panel Data Sample Selection Model», *Econometrica*, 65(6), 1335-64.
- [37] LANCASTER, T. (1979): «Econometric Methods for the Duration of Unemployment», *Econometrica*, 47(4), 939-56.
- [38] LANCASTER, T. (1990): *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.
- [39] MAGNAC, T. (1997): «State Dependence and Heterogeneity in Youth Employment Histories», *Working Paper*, INRA y CREST, París.
- [40] MANSKI, C. (1985): «Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator», *Journal of Econometrics*, 27, 313-33.
- [41] MANSKI, C. (1987): «Semiparametric Analysis of Ryom Effects Linear Models from Binary Panel Data», *Econometrica*, 55, 357-62.
- [42] McFADDEN, D. (1989): «A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration», *Econometrica*, 57, 995-1026.
- [43] MEYER, B. D. (1990): «Unemployment Insurance and Unemployment Spells», *Econometrica*, 58, 757-782.
- [44] MOON, C. G., y STOSTKY, J. G. (1993): «The Effect of Rent Control on Housing Quality Change: A Longitudinal Analysis», *Journal of Political Economy*, 101, 1114-48.
- [45] MUNDLAK, Y. (1978): «On the Pooling of Time Series y Cross-Section Data», *Econometrica*, 46, 69-85.
- [46] NARENDRANATHAN, W., y NICKELL, S. (1985): «Modelling the Process of Job Search», *Journal of Econometrics*, 28, 1, 29-49.
- [47] NARENDRANATHAN, W., y STEWART, M. (1993): «Modelling the Probability of Leaving Unemployment: Competing Risk Models with Flexible Baseline Hazards», *Applied Statistics*, 42, 1, 63-83.
- [48] NICKELL, S. (1979): «Estimating the Probability of Leaving Unemployment», *Econometrica*, 47 (5), 1249-66.
- [49] PAKES, A., y POLLARD, D. (1989): «Simulation and the Asymptotics of Optimization Estimators», *Econometrica*, 57, 1027-58.
- [50] PRENTICE, R. L., y GLOECKLER, L. A. (1978): «Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data», *Biometrics*, 34, 57-67.
- [51] RIDDER, G. (1987): «The Sensitivity of Duration Models to Misspecified Unobserved Heterogeneity and Duration Dependence», Mimeo, University of Amsterdam, Amsterdam.

- [52] SUEYOSHI, G. (1991): «Evaluating Simple Alternatives to the Proportional Hazards Models: Unemployment Insurance and the Duration of Unemployment», University of California, San Diego, febrero.
- [53] SUEYOSHI, G. (1995): «A Class of Binary Response Models for Grouped Duration Data», *Journal of Applied Econometrics*, 10, 411-431.